DOCUMENT RESUME

ED 407 167 PS 025 418

TITLE	Head Start: Research Provides Little Information on Impact of Current Program. United States General Accounting Office Report to the Chairman, Committee on the Budget, House of Representatives.
INSTITUTION	General Accounting Office, Washington, DC. Health, Education, and Human Services Div.
REPORT NO	GAO/HEHS-97-59
PUB DATE	Apr 97
NOTE	59p.
AVAILABLE FROM	U.S. General Accounting Office, P.O. Box 6015, Gaithersburg, MD 20884-6015; phone: 202-512-6000; fax: 301-258-4066; e-mail: info[at sign]www.gao.gov; world wide web home page: http://www.gao.gov (First copy, free; additional copies, $2 each. VISA and MasterCard credit cards are accepted).
PUB TYPE	Information Analyses (070) -- Reports - Evaluative (142) -- Reports - Research (143)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Accountability; Educational Quality; *Educational Research; Outcomes of Education; Preschool Education; *Program Effectiveness; Program Evaluation; Research Design; Research Methodology; *Research Needs; Research Problems
IDENTIFIERS	*Impact Studies; *Project Head Start

ABSTRACT

Although an extensive body of literature exists on Head Start, only a small part of this literature is program impact research. The body of existing research is inadequate in size, quality, and comparability for use in drawing conclusions about the impact of Head Start in any area in which Head Start provides services. This report from the General Accounting Office (GAO) provides an examination of the research plans of the Department of Health and Human Services (DHHS). Planned research will focus on new or innovative service delivery strategies (including descriptive studies, studies of program variations and demonstrations, and studies of the quality of innovative programs) but will provide little impact data on regular Head Start programs. DHHS believes effectiveness of Head Start is already proven, that conducting impact studies would be difficult and perhaps unethical, and that national estimates of program impact and comparisons with other service providers are not appropriate. GAO concluded that the increasing demand for shrinking federal resources and for accountability points to the need for conducting impact studies on Head Start and recommended that GHHS include in its research plan an assessment of the impact of regular Head Start programs. A summary of DHHS' comments conclude the report. Six appendices document how the GAO report was developed: Appendix 1, "Objectives, Scope, and Methodology"; Appendix 2, "Research, Demonstration, and Evaluation Budgets for the Head Start Program"; Appendix 3, "Summaries of Studies Included in the Review"; Appendix 4 "Comments from the Department of Health and Human Services"; Appendix 5 Acknowledgments"; and Appendix 6, "GAO Contacts and Staff Acknowledgments." (DR)

United States General Accounting Office

GAO

Report to the Chairman, Committee on the Budget, House of Representatives

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

April 1997

HEAD START

Research Provides Little Information on Impact of Current Program

ED 407 167

PS 025418

GAO/HEHS-97-59

BEST COPY AVAILABLE

GAO

United States
General Accounting Office
Washington, D.C. 20548

Health, Education, and
Human Services Division

B-271866

April 15, 1997

The Honorable John R. Kasich
Chairman, Committee on the Budget
House of Representatives

In the 30 years it has existed, Head Start has served over 15 million children at a total cost of $31 billion. Growing out of the War on Poverty in the mid-1960s, Head Start was created to provide comprehensive health, social, educational, and mental health services to disadvantaged preschool children. The program was built on the philosophy that effective intervention in the lives of children can best be accomplished through family and community involvement. Fundamental to program philosophy was the notion that communities be given considerable latitude to develop their own Head Start programs, an idea that has made variability a defining characteristic of the program. These philosophies and the general goals of the program remain virtually unchanged today.

Although Head Start has long enjoyed both congressional and public support, opinions about the program's impact have been divided and its effectiveness debated. Some maintain that no compelling evidence exists that Head Start makes any lasting difference in the lives of the population it serves. Others strongly support Head Start and maintain that research has conclusively established its value. Amid this debate, funding for Head Start has tripled in the past 10 years.

Conflicting information on program impact and the focus on results-oriented program performance information required by the Government Performance and Results Act (GPRA) of 1993 have renewed interest in the impact of the current Head Start program. In light of this, you asked us to determine (1) what the studies conducted on current Head Start programs[1] suggest about Head Start's impact and (2) what types of Head Start studies are planned by the Department of Health and Human Services (HHS).

We defined impact for this study as differences in outcomes caused by Head Start participation.[2] Implicit in this definition is the notion that

[1]Because Head Start changed significantly in its early years, which could have affected program impact, we defined current Head Start programs as those in existence in 1976 or later. See app. I for a further discussion of this decision.

[2]Our definition of impact is based on the concept of "net impact" as defined by the Office of Management and Budget (OMB) for agency use in developing performance measures for GPRA and is from "Selected Examples of Performance Measurement," OMB Office Memorandum 95-37 (Washington, D.C.: July 28, 1995).

differences found would not have occurred without program participation. Impact research is designed to permit the assumption that differences were caused by the program. Although impact studies are sometimes difficult and expensive, they are the only way to answer the question, "Is this program making a difference?" Thus, we included only studies in this review that gave some information on program impact. See appendix I for details on criteria we used to select studies.

To determine what research suggests about the impact of Head Start, we searched many electronic databases to locate published and unpublished manuscripts. We also spoke with early childhood researchers and practitioners to identify research studies. Our search yielded nearly 600 citations and documents, which were screened for possible inclusion in the study. Of these, we found 22 studies that fit our agreed-upon criteria and are reviewed in this study. (See app. I for details on our scope and methodology.) To obtain information about HHS' studies of Head Start, we reviewed HHS' research plans for Head Start and other research documents and spoke with HHS and National Head Start Association officials involved with Head Start research.

Results in Brief

Although an extensive body of literature exists on Head Start, only a small part of this literature is program impact research. This body of research is inadequate for use in drawing conclusions about the impact of the national program in any area in which Head Start provides services such as school readiness or health-related services. Not only is the total number of studies small, but most of the studies focus on cognitive outcomes, leaving such areas as nutrition and health-related outcomes almost completely unevaluated. Individually, the studies suffer to some extent from methodological and design weaknesses, such as noncomparability of comparison groups, which call into question the usefulness of their individual findings. In addition, no single study used a nationally representative sample so that findings could be generalized to the national program.

Failing to find impact information in existing research, we examined HHS' research plans for Head Start. Planned research will focus on new or innovative service delivery strategies and demonstrations but will provide little information on the impact of regular Head Start programs. HHS' planned research includes descriptive studies; studies of program variations, involving new and innovative service delivery strategies and demonstration projects; and studies of program quality.

HHS officials, in explaining the agency's research emphasis, stated that early research has proven Head Start's impact. Such research, however, conducted over 20 years ago, may no longer apply to today's program because of program changes and changes in the population served. HHS also noted some ethical and methodological difficulties of conducting impact research, especially studies that would produce national estimates of program effect. But neither ethical nor methodological issues present an insurmountable deterrent to conducting research on Head Start's impact. Moreover, the size and cost of the program appear to warrant an investment in such research.

Background

Begun in 1965 as a part of the effort to fight poverty, Head Start is the centerpiece of federal early childhood programs. Head Start's primary goal is to improve the social competence of children in low-income families, that is, their everyday effectiveness in dealing with both their present environment and later responsibilities in school and life. Social competence takes into account the interrelatedness of cognitive and intellectual development, physical and mental health, nutritional needs, and other factors. To support its social competence goal, Head Start has delivered a wide range of services to over 15 million children nationwide since its inception. These services consist of education and medical, dental, nutrition, mental health, and social services. Another essential part of every program is parental involvement in parent education, program planning, and operating activities.

Head Start services are provided at the local level by public and private nonprofit agencies that receive their funding directly from HHS. These include public and private school systems, community action agencies, government agencies, and Indian tribes. In fiscal year 1996, grants were awarded to about 1,400 local agencies, called grantees. Head Start grantees are typically required to obtain additional funding from nonfederal sources to cover 20 percent of the cost of their programs. The Head Start program works with various community sources to provide services. For example, some programs coordinate with public health agencies to obtain health services, while other programs contract with local physicians. Although all programs operate under a single set of performance standards, local programs have a great deal of discretion in how they meet their goals, resulting in great variability among programs.

Although the program is authorized to serve children at any age before the age of compulsory school attendance, most children enter the program at

age 4. The law requires Head Start to target children from poor families, and regulations require that 90 percent of the children enrolled in each program be low income. By law, certain amounts are set aside for specific subpopulations of children, including those with disabilities and Native American and migrant children.

In addition to providing services to children and families, Head Start also sees one of its roles as a national laboratory for child development. Consequently, Head Start uses much of its discretionary research funding for demonstrations and studies of program innovations. Although overall funding has grown over the years, the amount of funds allocated to research, demonstration, and evaluation[3] has represented about 2 percent or less of the Head Start budget. In fiscal year 1996, Head Start's research, demonstration, and evaluation budget totaled $12 million (see app. II).

Head Start Has Changed Over the Years

Today's Head Start is a much different program than it was 30 years ago. Although the program's goals have changed little since its inception, Head Start changed considerably during its first decade. Begun as a summer program, Head Start became largely a full-year program by the early 1970s. In addition, in the early to mid-1970s, the program launched improvement initiatives, including promulgation of performance standards and teacher credentialing. Programs also had the option of providing home-based services.

In the 1990s, the program continues to change. In 1990, the Congress passed the Head Start Expansion and Quality Improvement Act, which reauthorized Head Start and set aside funds for programs to use to enhance and strengthen the quality of services.[4] In 1994, the Congress established a new program—called Early Head Start—to serve low-income families with infants and toddlers. The program provides continuous, intensive, and comprehensive child development and family support services to low-income families with children under age 3.

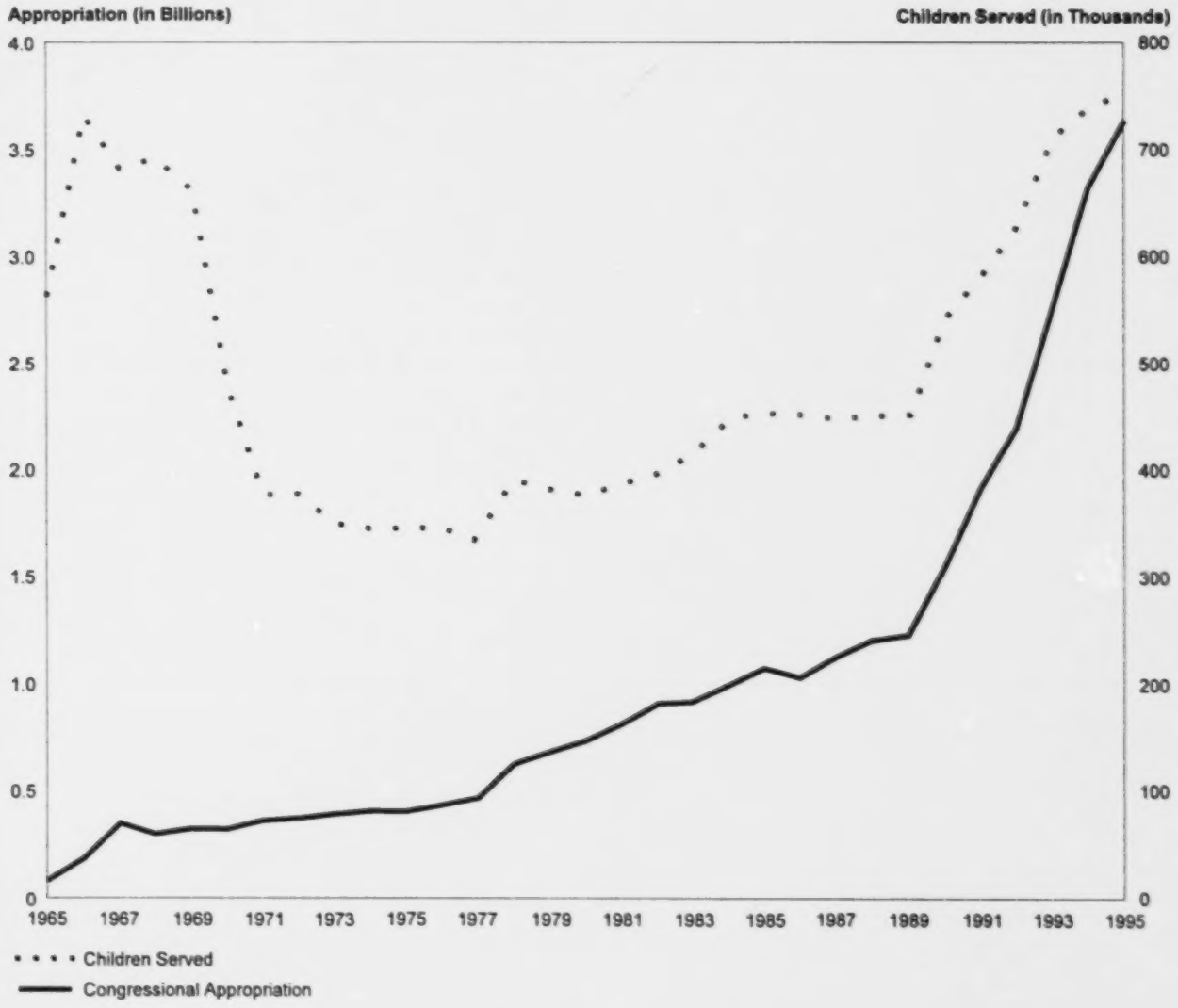
In addition to changes to Head Start over the years, other changes affecting the program relate to the children and families Head Start serves and the amount appropriated to support the program. Head Start's service population has become increasingly multicultural and multilingual and is confronted with difficult social problems such as domestic violence and

[3]HHS makes little distinction between spending for research and evaluation, HHS officials told us.

[4]Despite the emphasis on quality, some early childhood experts are still concerned about the uneven quality of Head Start programs.

drug abuse. Moreover, the number of children served by the program has grown dramatically—from 349,000 children in 1976 to about 750,000 in 1995. The amount appropriated for the program, which totaled $3.5 billion in 1995, has paralleled the growth in the number served (see fig. 1).

Figure 1: Growth in Head Start



Note: In the early years of Head Start, most programs were summer programs. In the early 1970s, summer programs were almost completely phased out.

Research on the Early Years of Head Start

In the decade after Head Start's inception, many studies of the program's impact were conducted. One of the first major studies was conducted for the Office of Economic Opportunity by the Westinghouse Corporation in 1969. This study found that summer Head Start programs produced no lasting gains in participants' cognitive or affective development and that full-year programs produced only marginal gains by grades one, two, and three. Several researchers criticized this study because of its methodology. Subsequently, many other studies investigated Head Start's impact.

In 1981, HHS contracted with CSR, Inc., to synthesize the findings of Head Start impact studies. CSR concluded that Head Start participants showed significant immediate gains in cognitive test scores, socioemotional test scores, and health status. Cognitive and socioemotional test scores of former Head Start students, however, did not remain superior in the long run to those of disadvantaged children who did not attend Head Start, according to CSR. In addition, on the basis of a small subset of studies, CSR reported that Head Start participants were less likely to be retained in grade and less likely to be placed in special education.[5]

Because these research studies were conducted during Head Start's infancy, their findings provide little information on the effectiveness of the current program. For instance, most of the programs included in the Westinghouse study were summer programs. Almost all programs today are full-year programs. Similarly, the great majority of studies in CSR's synthesis study were late 1960's and early 1970's programs and therefore would not have reflected many significant program changes that took place in the early to mid-1970s.

Interest in Impact Research Has Increased

Interest in Head Start's impact has grown with increased congressional and public concern for substantiating federal program performance. Traditionally, federal agencies have used the amount of money directed toward their programs, the level of staff deployed, or even the number of tasks completed as some of the measures of program performance. At a time when the value of many federal programs is undergoing intense public scrutiny, however, an agency that reports only these measures has not answered the defining question of whether these programs have produced real results. Because today's environment is results oriented, the Congress, executive branch, and the public are beginning to hold agencies

[5]Ruth McKey and others, The Impact of Head Start on Children, Families, and Communities, HHS Pub. No. (OHDS) 85-31193 (June 1985), p. 1.

accountable for outcomes, that is, program results as measured by the differences programs make.

The Congress' determination to hold agencies accountable for their performance lay at the heart of two landmark reforms of the 1990s: the Chief Financial Officers Act of 1990 and GPRA. With these two laws, the Congress imposed a new and more businesslike framework for management and accountability on federal agencies. In addition, GPRA created requirements for agencies to generate the information congressional and executive branch decisionmakers need in considering measures to improve government performance and reduce costs.

Body of Research on Current Head Start Program Insufficient to Draw Conclusions About Impact

The body of research on current Head Start is insufficient to draw conclusions about the impact of the national program. Drawing such conclusions from a body of research would require either (1) a sufficient number of reasonably well-designed individual studies whose findings could appropriately be combined to provide information about the impact of the national program or (2) at least one large-scale evaluation using a nationally representative sample. Findings from the individual studies we identified, however, could not be appropriately combined and generalized to estimate program impact at the national level. In addition, no single study used a nationally representative sample, permitting findings to be generalized to the national program.

Findings Could Not Be Combined to Produce National Estimates of Impact

The body of studies was inadequate to assess program impact by combining the findings of studies using similar outcome measures. The total number of studies found on Head Start impact was too small to permit generalizing findings to the national program. Most of these studies targeted cognitive outcomes, leaving other outcome areas, such as health and nutrition, scarcely examined. In addition, all the studies suffered to some extent from methodological problems that weakened our confidence in the findings of the individual studies.

Number of Studies Too Small

Although the body of literature on Head Start is extensive, the number of impact studies was insufficient to allow us to draw conclusions about the impact of the national Head Start program. Such an aggregation of findings should be based on a large number of studies.[6] The larger the number of studies, the greater the chance that the variability in Head Start programs

[6]John E. Hunter and Frank L. Schmidt, Methods of Meta-Analysis (Newbury Park, Cal.: Sage Publications, 1990), p. 83.

would be represented in the studies. Conversely, the smaller the number of studies, the greater the risk that the aggregate findings from these studies may not apply to Head Start in general. Most of the approximately 600 articles and manuscripts about Head Start that we identified could not be used to answer questions about impact for various reasons. Much of this literature consisted of program descriptions, anecdotal reports, and position papers. Of those articles that were research studies, some (for example, case studies) were not suitable for drawing general conclusions about impact. Some studies examined change in outcome measures before and after Head Start but did not control for other plausible explanations for the change, for example, maturation. Other studies using a comparison group to control for competing explanations of change did not provide statistical information about the confidence that the differences found were not chance occurrences.

Only 22 of the more than 200 manuscripts we reviewed met our criteria for inclusion in our analysis. (See app. I for a detailed description of inclusion criteria.) Of these, 16 investigated impact by comparing Head Start participants with an unserved comparison group; 3 analyzed gains on normed tests. Only three studies included comparisons of Head Start with some other type of preschool or day care program. These studies represent work by a variety of researchers, including college students, college faculty, and contractors. Appendix III contains more detailed information on each study.

No Outcome Area or Population Adequately Researched

Although Head Start provides services in several outcome areas, such as health, nutrition, education, and the like, most of the studies we found focused on educational/cognitive outcomes, and few made distinctions on the basis of differing populations served by Head Start. For example, most of the studies examined the impact of Head Start on grade retention and other indicators of academic achievement, such as standardized reading and math scores. Of the 22 studies included in our review, 16 included one or more outcomes in the cognitive area. Conversely, only five studies investigated health- or nutrition-related outcomes, and only five examined family impacts.

Similarly, few studies analyzed impact by subpopulations. Because Head Start is a multicultural program, serving children and families of varying races, ethnic backgrounds, and socioeconomic levels, research that targets these subpopulations may uncover differential effects.

Studies Suffered From Methodological Weaknesses

All of the studies had some methodological problems. Although research in field settings can rarely conform to rigorous scientific procedures, in general, researchers place more confidence in findings of studies that control for competing explanations for their results and that use large samples.

One of the more serious of the methodological problems was noncomparability of comparison groups. The most reliable way to determine program impact is to compare a group of Head Start participants with an equivalent group of nonparticipants. The preferred method for establishing that the groups are equivalent at outset is to randomly assign participants to either the Head Start group or the comparison group. Only one of the studies we reviewed used random assignment to form the Head Start and non-Head Start comparison groups. Most of these studies formed a comparison group by selecting children who were similar to the Head Start participants on some characteristic thought to be important to the outcome under study. In most cases, researchers matched participants on one or more demographic variables, usually including some variable related to socioeconomic level. In other cases, researchers did not match treatment and comparison groups but tried to compensate statistically for any inequality between the groups. Neither of these methods compensates completely for lack of random assignment to group.

Some of the studies used no comparison group; instead, they compared performance of Head Start participants with test norms. This approach to evaluating program performance indicates the performance of Head Start participants relative to the norming group. Because the norming group may be unlike the Head Start group, however, conclusions about program impact are unclear.

Finally, many of the studies also suffered from small samples, especially those investigating intermediate and long-term effects. Some studies began with relatively small samples; others, which began with larger samples, ended up with smaller samples as the study progressed because of missing data and attrition. Small samples present problems in research because they adversely affect statistical procedures used in analyses. Some procedures cannot appropriately be used with small samples; others are rendered less able to detect differences, resulting in an underestimation of program effects.

No National Program Evaluation Found

No completed, large-scale evaluation of any outcome of Head Start that used a nationally representative sample was found in our review. One characteristic of Head Start is program variability, not only in the kind of services delivered, but also in the quality of services. Making summary statements about program impact requires that the sample of programs studied represent all programs nationwide.

Although one evaluation had a study design that would have allowed findings to be generalized to the national program, this study was never completed. In the late 1970s, HHS contracted for a national evaluation of the educational services component of basic Head Start. The design called for a longitudinal study that would follow children and their parents from preschool through the fourth grade. The evaluation was to compare the Basic Educational Skills Program, regular Head Start, and a non-Head Start control group. Thirty Head Start programs were to be randomly selected, and Head Start-eligible children from these communities were to be randomly assigned to Head Start or the control group. Many methodological problems as well as funding problems occurred, however, during the implementation of this study, and it was abandoned.

The 1990 act that reauthorized funding for Head Start directed the Secretary of HHS to conduct ". . . a longitudinal study of the effects that the participation in Head Start programs has on the development of participants and their families and the manner in which such effects are achieved." The study, as described in the act, was to examine a wide range of Head Start outcomes, including social, physical, and academic development, and follow participants at least through high school. The description also stipulated that, "To the maximum extent feasible, the study . . . shall provide for comparisons with appropriate groups composed of individuals who do not participate in Head Start programs." The act authorized the appropriation of funds to carry out this study for fiscal years 1991 through 1996. According to HHS, however, funds were never appropriated for the study, and it was not conducted.

Research Planned by HHS Focuses on Program Improvement, Not Impact

Head Start's planned research will provide little information about the impact of regular Head Start programs[7] because it focuses on descriptive studies; studies of program variations, involving new and innovative service delivery strategies and demonstration projects; and studies of program quality. Although these types of studies are useful in evaluating programs, they do not provide the impact information needed in today's results-oriented environment and encouraged by GPRA.

HHS Focuses Research on Program Improvement

The primary focus of research, according to Head Start Bureau officials, is to improve the program by exploring ways to maximize and sustain Head Start benefits. Thus, HHS studies evaluate which practices seem to work best for the varying populations Head Start serves and ways to sustain program benefits. Some of these studies are descriptive, providing information on service delivery and the characteristics of populations receiving services. For example, HHS is currently conducting a descriptive study of the characteristics of families served by the Head Start Migrant Program. Other descriptive studies have been conducted on health services and bilingual/multicultural programs.

HHS also funds studies designed to answer questions about the effectiveness of new or innovative service delivery strategies and demonstrations and how effectiveness may relate to characteristics of the population served. Such studies typically involve special program efforts and demonstration projects conducted on a trial basis at a few Head Start sites that focus on practices or services not typically found in regular Head Start programs.[8] For example, both Early Head Start and the Comprehensive Child Development Program target infants and children younger than those normally served by Head Start. Similarly, the Family Service Center demonstrations place more emphasis on family services and provide assistance in a variety of areas such as illiteracy, substance abuse, and unemployment.

In addition, HHS funds research to explore program quality and to develop instruments to assess program performance. In 1995-96, HHS funded several Quality Research Centers and a Performance Measure Center to

[7]"Regular" Head Start refers in this report to programs that operate within the scope of established Head Start program options and under normal Head Start requirements. Regular programs are to be distinguished from demonstrations and other special programs, which may serve populations or offer services not normally found in Head Start.

[8]Both special programs and demonstrations are innovative programs, implemented on a limited basis and with program features not found in regular Head Start programs. Demonstrations, however, have a predetermined end because their grants expire at the end of a specified period. Special programs may continue to receive funding because programs may recompete for such grants at the end of the grant period.

develop and identify instruments for measuring the quality of Head Start programs and to collect performance measure data on a nationally representative sample of Head Start programs. The major purpose of this effort, according to HHS officials, is to determine which program characteristics relate to meeting program goals. Some of the performance measure assessments use instruments for which national norms are available, however, and HHS will be able to compare participant performance to national norms for these measures.

Identifying performance measures is an important step in building a research and evaluation base for Head Start. Because the program's goals are so broad and difficult to assess, precisely defining expected outcomes and identifying appropriate instruments should produce a more valid, useful body of research. But identifying standard performance measures is also valuable because it provides a set of common measures upon which a body of research could be built, including impact research.

Although descriptive studies, studies of new or innovative programs and demonstrations, and studies of program quality provide information useful both to HHS and the Congress, they do not provide full information on the impact of regular Head Start. Even the performance measures study already discussed will not provide clear-cut impact information because no comparison group is being used. Over time, this type of study will provide some useful information about program outcomes; however, such a study can neither attribute effect nor estimate the precise effect size with the level of confidence found in comparison group studies.

Research Planned by HHS Will Provide Little Information on Program Impact

Research planned by HHS will provide little program impact information on regular Head Start programs. HHS officials expressed concerns about using their research dollars for impact research rather than program improvement. The effectiveness of Head Start has been proven by early research, according to these officials, who also pointed to difficulties in conducting impact studies. In addition, because Head Start is such a varied program, averaging across local programs to produce national estimates of effect is not appropriate, they said. Finally, HHS maintains that Head Start is unique because of the comprehensiveness of services it offers and the population it serves; therefore, comparing Head Start with other service programs would be inappropriate, HHS officials believe.

Most of the research that HHS cited as evidence of Head Start's impact is outdated, however, and, as previously mentioned, insufficient research has

been done in the past 20 years to support drawing conclusions about the current program. Furthermore, it appears that impact studies on Head Start could be done and would provide valuable results-oriented information. In addition, although research on programs that vary greatly could be methodologically more challenging to producing national estimates of impact, variation alone should not prevent developing such estimates. Moreover, comparisons with other service programs, if designed to answer questions about specific program outcomes, would provide useful information about assessing program impacts.

HHS Believes Effectiveness of Head Start Is Already Proven, So Further Impact Research Is Not Warranted

HHS maintains that early research has proven the effectiveness of early childhood education, including Head Start, so impact research is not the most effective use of limited research funds. Findings from early studies, however, do not conclusively establish the impact of the current Head Start program because today's program differs from that of the late 1960s and early 1970s. Although program changes might be assumed to increase positive impact, this assumption is largely unsubstantiated. In addition, program impact may be affected by changes in the population served; Head Start families today face different problems than those in the past because of an increase in substance abuse, violence, and homelessness. Furthermore, an increased availability of social services may have lessened the impact of Head Start because families may get services from other sources if not from Head Start. The net effect of these changes on program impact is unknown.

Later studies offered to support Head Start's impact do not provide enough evidence to conclude that current Head Start is effective. Findings in literature reviews cited by Head Start proponents to support its effectiveness often involve only a few Head Start programs. For example, HHS cited a review in a recent Packard Foundation report[9] that reported positive cognitive results of early childhood programs. This review, however, had only five studies involving Head Start participation in 1976 or later, and two of the five studies combined Head Start and other public preschools in the analyses. Authors of other studies of high-quality preschool programs have sometimes warned against applying their findings to Head Start. For instance, researchers in the Consortium for Longitudinal Studies, which produced a major study reporting positive long-term effects of preschool, explicitly stated that caution should be used in generalizing their findings to Head Start and that the programs

[9]Steven W. Barnett, "Long-Term Effects of Early Childhood Programs on Cognitive and School Outcomes," in The Future of Children: Long-Term Outcomes of Early Childhood Programs, Richard E. Behrman, ed., Vol. 5, No. 3 (Los Altos, Cal.: 1995).

were ". . . examples of what Head Start could be rather than what it has been."[10]

HHS Believes Conducting Impact Studies Would Be Difficult

HHS believes conducting impact research would present methodological difficulties. Two types of research designs are commonly used in conducting impact studies, experimental and quasi-experimental. HHS officials mentioned difficulties with both types of designs in studying Head Start's impact. In addition, finding enough unserved children to form comparison groups would be a problem with either kind of research design, they said.

True experimental designs, also called randomized trials, are comparison group studies that randomly assign study participants to either a treatment or control group. In the case of Head Start, these studies would require recruiting more eligible children than the program can serve. From these recruits, some children would be randomly assigned to Head Start; the rest, the unserved children, would constitute the control group. HHS officials cited ethical considerations of assigning children to an unserved control group as one of the difficulties in conducting randomized trials.

Randomized trials, however, could be appropriately applied to Head Start research. In fact, the evaluation of the Early Head Start project, now under way, has randomly assigned potential participants to Early Head Start or a control group that has not received Early Head Start services. Alternatively, a research design that delays, rather than withholds, services could be used. This would involve selecting a study group and randomly assigning some children to Head Start the first year, while the remainder would serve as a control group. The control group would receive services the following year. Another strategy that could be used to study specific parts of the program would be to use an alternative treatment design. In this case, some randomly assigned participants would receive the full Head Start program, while others would receive partial services. For example, if the study interest is in school readiness and cognitive issues, the control group might receive only nutritional and health services.

Most researchers believe that randomized trials yield the most certain information about program impact. Random assignment is an accepted practice in virtually every area of research, including medicine, economics, and social sciences. In some cases, the treatment of study interest is simply withheld from the control group. In other cases, for

[10]Sandra Condry, "History and Background of Preschool Intervention Programs and the Consortium for Longitudinal Studies," in As the Twig Is Bent . . . Lasting Effects of Preschool Programs, Lawrence Erlbaum Associates (Hillsdale, N.J.: 1983), p. 27.

example, when researchers suspect that withholding treatment would have a profoundly negative impact, treatment may be delayed for awhile or some lesser, alternative treatment offered. While acknowledging the difficulties of random assignment, some early childhood researchers we spoke with suggested that Head Start conduct randomized trials to study regular Head Start programs because this type of study provides the most conclusive information on program impact.

A common alternative to randomized trials, quasi-experimental designs, uses a naturally occurring, unserved comparison group. In the case of Head Start, some researchers have tried to identify other children in the community who are like Head Start participants in ways thought to be important (usually socioeconomic level) but who are not enrolled in Head Start. This group became the comparison (control) group.

Quasi-experimental research is less rigorous than research that uses random assignment, and less confidence can be placed in its conclusions. Rarely are pre-existing groups equivalent. Even when statistical adjustments are made to compensate for known nonequivalencies, some questions always remain about the degree to which pre-existing differences in the groups may have contributed to study results. When well planned and well executed, however, such designs can provide some indication of program impact.

Because Head Start strives to serve the neediest children, those in quasi-experimental comparison groups would be less likely to be disadvantaged than children in the Head Start group, according to HHS officials. If true, this nonequivalency in groups would bias the outcome in favor of the comparison group, resulting in underestimation of program effects. Because investigating the characteristics of Head Start participants was beyond the scope of this study, we do not know to what extent, if any, Head Start children may be more disadvantaged than similar children not attending Head Start. Even assuming that Head Start has identified and is serving the neediest applicants, however, it seems possible that a comparably disadvantaged, unserved group could be identified from the applicants whom the program cannot serve and nonapplicants in a community.

Regardless of which design is used, experimental or quasi-experimental, finding enough truly unserved children for a comparison group would be extremely difficult because of the growing number of public preschool programs and the increased availability of child care, according to HHS

officials. Statistics on the percentage of children being served by preschools suggest, however, that finding disadvantaged children unserved by preschools is possible. In our report, Early Childhood Programs: Many Poor Children and Strained Resources Challenge Head Start (GAO/HEHS-94-169BR), we found that only 35 percent of poor 3- and 4-year-olds attended preschool in 1990. The Congressional Research Service estimated that in fiscal year 1994, about 30 percent of eligible 3- to 4-year-olds were being served by Head Start.[11] On the basis of these estimates, it appears that some locations do exist where a control group of children not attending preschool could be formed.

HHS Believes National Estimates of Program Impact Are Not Appropriate

Estimating program impact at the national level is not appropriate because of the extreme variability of local programs, HHS officials said. Local Head Start sites have great flexibility, and, even though all programs share common goals, they may operate very differently. Therefore, on the advice of HHS' research advisory panel,[12] HHS considers a single, large-scale, national study of impact to be methodologically inappropriate. For this same reason, HHS believes that summing across sites for an aggregate estimate of effect is not justified in cases where sites are not basically operating the same way.

Evaluating outcomes at the national program level is an accepted program evaluation procedure, however, even for programs with a great deal of variability. It is the only way to determine with certainty whether the program is making an overall difference in any particular outcome area. Aggregate analysis does not, however, replace the need for lower level analyses, which provide insight into the summary finding. In cases where effects are not uniform across sites, this lower level analysis provides more understanding of which service areas and delivery approaches are working for which subpopulations. Evaluations can be planned to answer both the aggregate and disaggregate question in a single study.

HHS Believes Comparisons With Other Service Providers Are Not Appropriate

Another way to evaluate Head Start's impact is to compare its effects with some other types of preschool, for instance, state or local preschools. When several programs exist that deliver similar services, studies comparing programs in areas that have common goals can provide useful information. For instance, Head Start and public preschools share the goal of school readiness. A study might be conducted to compare Head Start

[11]Head Start in the 104th Congress, Congressional Research Service Report for Congress (Washington, D.C.: Mar. 6, 1996), p. 2.

[12]Since 1989, advisory panels of experts in early childhood education and research and evaluation have helped HHS plan its research.

and public preschool students on the basis of a measure of school readiness. Such a study might compare the performance of program participants, while describing relevant program differences that might affect results, such as level of service in the area studied and program costs.

Regarding a comparative study, HHS has maintained that Head Start is unique in the comprehensiveness of the services it offers. Therefore, according to the agency, any comparison of programs would be misleading. In addition, HHS claims that children served by Head Start are more disadvantaged than children in other types of preschools. The agency also points out that in some places, other public preschools have adopted the Head Start model, making such comparisons essentially Head Start with Head Start.

Concerns about differences in populations served by the programs would relate to the rigor of the study design, that is, whether it is experimental or quasi-experimental. When quasi-experimental designs are used, researchers frequently use statistical techniques to mitigate for pre-existing differences; but these designs always suffer to some degree from the limitations referred to earlier in our discussion of quasi-experimental designs. Therefore, confidence in the study's results would vary depending on the study design used.

In the case of Head Start-like programs, one might reasonably expect a difference in outcome on the basis of such factors as program administration and context. For example, a preschool program operated by a local school system might have different outcomes in school readiness because of the possible advantage of transitioning its students into kindergarten. Research that compares Head Start with alternative ways of accomplishing a particular goal might provide insight into the most effective and efficient way to provide services to needy children and families.

Conclusions and Recommendations

Increasing demand for shrinking federal resources has raised the concerns of the Congress, the executive branch, and taxpayers about the impact of multibillion dollar federal investments in federal programs such as Head Start. In addition, GPRA requires agencies to be more accountable for substantiating program results. Although research has been conducted, it does not provide information on whether today's Head Start is making a

positive difference in the lives of participants who live in a society that differs vastly from that of the sixties and early seventies.

While we acknowledge the difficulties of conducting impact studies of programs such as Head Start, research could be done that would allow the Congress and HHS officials to know with more certainty whether the $4 billion dollar federal investment in Head Start is making a difference. For this reason, we recommend that the Secretary of HHS include in HHS' research plan an assessment of the impact of regular Head Start programs.

Agency Comments

In commenting on a draft of our report, HHS expressed the belief that the research base on the efficacy of Head Start is more substantial than depicted in our report and that the strategy of the Department to extend this base is appropriate to produce findings about both impact and program quality. HHS also indicated plans to evaluate the feasibility of conducting impact studies such as we recommended. The Quality Research Centers are evaluating the feasibility of conducting randomized trials in small-scale evaluations, and, on the basis of these experiences may consider implementing larger scale studies. The full text of HHS' comments appears in appendix IV.

HHS supported the claim that the research base is more substantial than we depict by pointing to the findings from the 1985 synthesis conducted by CSR (cited as "McKey et al., 1985" in HHS' comments) and two more recent studies (the Currie and Thomas study and the Fosburg study). For reasons discussed in this report, we do not agree that findings drawn from studies more than 20 years old adequately support claims about the impact of the current Head Start program. Similarly, the findings from the two more recent studies mentioned fail to support conclusions about impact that can be generalized to the national program. Even though these studies were larger than others we found, both had significant methodological limitations. The Currie and Thomas study examined information in a database to reach conclusions about Head Start. This study used an after-the-fact, post-test-only design. Although this design is frequently used when researchers must rely on existing data as their only source, the design is vulnerable to serious threats to validity, as discussed earlier in this report. Because of these design limitations, neither positive conclusions about Head Start (that is, that children's test scores show immediate positive effects) nor negative conclusions (that is, that these effects quickly disappear for African American children) can be firmly drawn from the findings of this study.

The second study, the Fosburg study, as HHS pointed out, used a much stronger research design, which randomly assigned children in four Head Start programs to either Head Start or a non-Head Start control group. The site selection methodology, however, precluded generalizing these findings to all Head Start programs. The four Head Start programs selected were chosen from areas identified as underserved in medical and dental services, and Head Start sites that were not in compliance with Head Start performance standards were excluded from selection. In addition, attrition was a significant problem in this study.

HHS also mentioned that on the basis of recommendations of leading researchers, the Department is conducting a well-balanced, innovative set of new studies of Head Start. It contends that our report does not acknowledge the major longitudinal studies that HHS has planned or that are being conducted by other agencies. Our report states that HHS' planned research focuses on program improvement, and we agree that such studies are needed. We also support the studies of program impact that HHS has under way in special program areas such as Early Head Start. Our work, however, focused specifically on HHS' research plans that address the question of impact of the regular Head Start program. HHS' current research plans, however, do not include such research.

Finally, HHS maintained that it is building a substantial system of innovative research, development, and management tools in response to GPRA. The Department emphasized the role the Quality Centers play in these efforts and said that these centers are currently evaluating possible strategies for performing comparison group studies that use a random assignment research design. HHS maintained that we overlooked the importance of studying the quality of Head Start programs in assessing impacts.

We fully support HHS' plans to investigate the feasibility of conducting randomized trials because these studies provide the clearest indication of program impact. We also agree that the issue of quality is important in assessing program impact and findings from studies need to include information on program quality. The ultimate measure of program quality is impact, however. Until sound impact studies are conducted on the current Head Start program, fundamental questions about program quality will remain.

We are sending copies of this report to the Secretary of Health and Human Services, the Head Start Bureau, appropriate congressional committees, the Executive Director of the National Head Start Association, and other interested parties. Please call me at (202) 512-7014 if you or your staff have any questions about this report. Major contributors to this report are listed in appendix VI.

Sincerely yours,

Carlotta C. Joyner
Director, Education and
Employment Issues

Contents

Letter	1
Appendix I Objectives, Scope, and Methodology	24
Appendix II Research, Demonstration, and Evaluation Budgets for the Head Start Program	27
Appendix III Summaries of Studies Included in the Review	28
Appendix IV Comments From the Department of Health and Human Services	47
Appendix V Acknowledgments	53
Appendix VI GAO Contacts and Staff Acknowledgments	54
Related GAO Products	56

Figure

Figure 1: Growth in Head Start 6

Abbreviations

GPRA	Government Performance and Results Act
HHS	Department of Health and Human Services
PPVT-R	Peabody Picture Vocabulary Test-Revised
WJ-R	Woodcock-Johnson Tests of Achievement

Objectives, Scope, and Methodology

Objectives

The Chairman of the Committee on the Budget, House of Representatives, asked us to examine existing research on Head Start programs and to determine what it suggests about the impact of the current Head Start program. Another objective was to determine what types of Head Start research HHS has planned.

Scope

Although the bulk of research on Head Start was conducted in the early years of the program, we focused on studies of Head Start participation in 1976 or later for several reasons. First, HHS instituted quality initiatives and other important program changes in the early to mid-1970s that shaped the current Head Start program, including phasing out summer programs, implementing performance standards, and establishing teacher credentialing procedures. Second, findings from studies of early programs have limited generalizability to more stable programs. The early years of any program are not likely to represent a program in its maturity. This is especially true for Head Start, which was implemented quickly and on a large scale.

Finally, earlier studies were thoroughly reviewed by the Head Start synthesis project and were reported in The Impact of Head Start on Children, Families, and Communities in 1985. Studies from the years before 1976 constituted the bulk of studies included in this synthesis. A short summary of these findings appears in the "Background" section of this report.

After speaking with HHS research personnel, we anticipated that the body of studies usable for a research synthesis might be small. Therefore, in addition to comparison group studies, we included pretest/post-test-only designs in cases in which outcomes were discussed in relation to test norms. Although much less useful in providing information about program impact relative to comparison group designs, these studies provide a certain degree of valuable information.

Methodology

To report on what existing research says about Head Start's impact, we identified studies meeting our basic selection criteria as outlined in the "Literature Review" section of this appendix. Because the number of studies found in the first phase was so small, we did not screen further for adequacy of information reported. To determine how HHS uses research, we reviewed HHS' research plans and publications by their research advisory panel. We also spoke with HHS officials who direct Head Start

research and with the director of research at the National Head Start Association.

Literature Search

We began our search for studies with two bibliographies contracted for by HHS. The first, An Annotated Bibliography of the Head Start Research Since 1965, was a product of the 1985 Head Start Evaluation, Synthesis and Utilization Project. We also reviewed An Annotated Bibliography of Head Start Research: 1985-1995. This bibliography was produced by Ellsworth Associates, Inc., for the Head Start Bureau as a part of its contract to maintain a library of Head Start and related research. Search strategies used to compile these bibliographies are described in the introductions to the documents.

In addition, we conducted our own search for studies. Our primary source was the database maintained by the Education Resources Information Center. However, we also searched a number of other databases, including MEDLINE, AGRICOLA, Dissertation Abstracts, Government Printing Office, Mental Health Abstracts, Psyc INFO, Federal Research in Progress, Social SciSearch, Sociological Abstracts, IAC Business A.R.T.S., British Education Index, Public Affairs Information Service International, and National Technical Information Service.

We also interviewed people knowledgeable about early childhood research. We attended the Head Start Third National Research Conference and spoke with conference participants. We also mailed letters to every conference participant asking for their assistance in locating relevant research. We interviewed personnel in charge of research for the Administration for Children Youth and Families and the Head Start Bureau and spoke with other researchers whom they recommended. We also talked with the executive director and the director of research and evaluation of the National Head Start Association and addressed the state and regional presidents of this organization at their annual meeting. In addition, we announced our effort to locate research dealing with Head Start effectiveness on several of the Internet forums sponsored by the American Educational Research Association.

Literature Review

From these sources, we identified over 600 manuscripts that were screened for relevance to our study. We acquired about 200 of these and reviewed them carefully regarding the following selection criteria:

- Head Start participation had occurred in 1976 or later;
- studies had compared outcomes of Head Start participants with children not attending any preschool—or those attending some other type of preschool—or studies had compared Head Start outcomes with test norms; and
- tests of statistical significance[13] were reported to have been performed on the differences, except in cases in which outcomes were measured using normed instruments.

We excluded studies of transition or follow-through programs that provided services beyond the Head Start years and studies that pooled Head Start and other kinds of preschool participants. We considered multiple articles or later follow-ups on the same study to be one study. This final screening yielded 22 impact studies that were evaluated in our review.

We performed our work between April and December 1996 in accordance with generally accepted government auditing standards.

[13]Tests of statistical significance produce a measure of the likelihood that a finding occurred as a result of sampling error. Such tests result in a value (usually denoted by "p") that represents the probability that the outcome (for example, differences in scores) arose from a sampling error. Thus, "p< .01" means that the probability of the outcome occurring as a result of sampling variation is less than 1 in 100. Although statistical significance may take on any value from .00 to 1.0, benchmark levels often used in research are .01 and .05. We excluded one study because of lack of evidence about the statistical significance of the findings.

Research, Demonstration, and Evaluation Budgets for the Head Start Program

Fiscal year	Appropriation	Research, demonstration, evaluation amount	Percentage of total appropriation
1971	$360,000,000	$5,700,000	1.58
1972	376,300,000	6,200,000	1.65
1973	400,700,000	9,000,000	2.25
1974	403,900,000	9,000,000	2.23
1975	403,900,000	9,000,000	2.23
1976	441,000,000	9,000,000	2.04
1977	475,000,000	9,000,000	1.89
1978	625,000,000	8,200,000	1.31
1979	680,000,000	14,630,000	2.15
1980	735,000,000	15,200,000	2.07
1981	818,700,000	14,600,000	1.78
1982	911,700,000	12,300,000	1.35
1983	912,000,000	6,900,000	0.76
1984	995,750,000	1,800,000	0.18
1985	1,075,059,000	1,300,000	0.12
1986	1,040,315,000	810,000	0.08
1987	1,130,542,000	1,300,000	0.11
1988	1,206,324,000	1,300,000	0.11
1989	1,235,000,000	1,500,000	0.12
1990	1,552,000,000	1,500,000	0.10
1991	1,951,800,000	3,500,000	0.18
1992	2,201,800,000	8,500,000	0.39
1993	2,776,286,000	8,900,000	0.32
1994	3,325,728,000	12,000,000	0.36
1995	3,534,128,000	12,000,000	0.34

Summaries of Studies Included in the Review

Evaluation of the Process of Mainstreaming Handicapped Children Into Project Head Start, Phase II, Executive Summary, and Follow-Up Evaluation of the Effects of Mainstreaming Handicapped Children in Head Start

Authors: Applied Management Sciences, Inc. (first study) and Roy Littlejohn Associates, Inc. (second study)

Outcome area studied: Cognitive and health

Overview of study: Children receiving Head Start program services compared with children receiving services from other types of programs and with children receiving no special services

Design: Pretest with post-test 6 months later

Population: 55 randomly selected Head Start centers and 49 non-Head Start programs

Sample: 391 Head Start children, 321 non-Head Start children, and 121 unserved children

Head Start program year(s): 1977-78

Measures/instrumentation: Various development indicators, including physical development, self-help skills, cognitive development, social development, communication skills, classroom social skills, and classroom behavior and social integration

Findings: Developmental gains for Head Start and non-Head Start children identified as physically handicapped, mentally retarded, and health or developmentally impaired were generally not significantly greater than those of unserved children. Developmental gains were significant in physical, self-help, academic, and communications skills for children identified as speech impaired in Head Start and non-Head Start programs relative to unserved children.

Source: Applied Management Services, Inc., Evaluation of the Process of Mainstreaming Handicapped Children Into Project Head Start, Phase II, Executive Summary, Education Resources Information Center (ED 168291), 1978, and Roy Littlejohn Associates, Inc., Follow-Up Evaluation of the Effects of Mainstreaming Handicapped Children in Head Start (Washington, D.C.: 1985).

A Longitudinal Study to Determine If Head Start Has Lasting Effects on School Achievement

Author: Colleen K. Bee

Outcome area studied: Cognitive

Overview of study: Followed up Head Start participants in kindergarten, first grade, and second grade

Design: Post-test only, comparison group selected from waiting list for each respective year

Population: Head Start participants in Sioux Falls, South Dakota

Sample: 10 girls and 10 boys were selected for each Head Start year, 10 girls and 10 boys were selected each year for the comparison groups

Head Start program year(s): 1977-78, 1978-79, 1979-80

Measures/instrumentation: Metropolitan Reading Readiness Test, special education placements, and grade retention

Findings: No significant differences were found at the .01 level of confidence on reading readiness scores for any of the years studied. Non-Head Start group retained in grade less than the Head Start group in 1977-78 (difference significant at .01 level). No significant difference was found in special education placements for any of the years studied.

Source: Colleen K. Bee, A Longitudinal Study to Determine If Head Start Has Lasting Effects on School Achievement, UMI Dissertation Services (Ann Arbor, Mich.: 1981).

Evaluation of Public Preschool Programs in North Carolina

Authors: Donna M. Bryant, Ellen S. Peisner-Feinberg, and Richard M. Clifford

Outcome area studied: Cognitive and socioemotional

Overview of study: Followed up public preschool graduates in kindergarten

Design: Post-test only, comparison group comprised children from same kindergarten classes

Population: Public preschool programs in North Carolina

Sample: 97 children participated in Head Start, 99 in community day care, and 120 in no group care

Head Start program year(s): 1992-93

Measures/instrumentation: Reading and math subscales of the Woodcook-Johnson Tests of Achievement (WJ-R), Peabody Picture Vocabulary Test-Revised (PPVT-R), developmental assessment on communication development, and an assessment on social behavior completed by a kindergarten teacher; adapted questionnaire form of the Communication Domain of the Vineland Adaptive Behavior Scale used to provide a measure of children's cognitive development; and Social Skills Questionnaire used to measure teachers' ratings of children's classroom behaviors

Findings: Significant group effects were found for the PPVT-R, with all the groups performing better than children in the non-day care group. For the WJ-R reading scale, the preschool group showed no effects. Significant main effects were found for preschool group on the WJ-R math scale, with the community day care sample scoring higher than the four other groups. Significant preschool group differences were found on the Vineland Communication Domain, with community day care children rated higher than the other four groups. Social skills of community child care children were rated significantly higher by their kindergarten teachers than children who attended the standard or the family-focused classes or children who did not attend group day care and marginally higher than the Head Start children. On the Academic Competence scale of the Social Skills Questionnaire, children who previously attended community child care scored significantly higher than those in the other four groups.

Source: Donna M. Bryant, Ellen S. Peisnar-Feinberg, and Richard M. Clifford, Evaluation of Public Preschool Programs in North Carolina, University of North Carolina (Chapel Hill, N.C.: 1993).

The Impact of Escalating Family Stress on the Effectiveness of Head Start Intervention

Authors: Mary Anne Chalkley and Robert K. Leik

Outcome area studied: Family

Overview of study: Explored the effects that declining conditions among the U.S. poor may have on the potential for intervention programs to make a difference in the lives of those receiving services

Design: Pretest/post-test, followed up Head Start Family Impact Project participants in 1993; comparison group recruited from Head Start-eligible families

Population: Head Start families in Minneapolis, Minnesota

Sample: 130 of the 190 families in the original study

Head Start program year(s): 1986-87, 1989-90

Measures/instrumentation: Mothers reported various measures on their families, themselves, and their children. Children completed the pictorial form of Perceived Competence and Acceptance.

Findings: An examination of the absolute amount of change in the mother's perception of the child was inconclusive on the impact of Head Start.

Source: Mary Anne Chalkey and Robert K. Leik, The Impact of Escalating Family Stress on the Effectiveness of Head Start Intervention, paper presented at the National Head Start Association's 22nd Annual Training Conference (Washington, D.C.: April 1995).

Developmental Progress of Children Enrolled in Oklahoma Head Start Programs in 1987-1988

Author: Laurna Champ

Outcome area studied: Cognitive and social

Overview of study: Head Start students were tested in the fall and again in the spring in multiple developmental areas.

Design: Pretest/post-test

Population: Children in 15 Head Start programs in Oklahoma

Sample: 120 students

Head Start program year(s): 1987-88

Measures/instrumentation: Brigance Diagnostic Inventory of Early Development and Head Start Measures Battery

Findings: Gains on the Brigance ranged from 9 to 16 months. Similar claims were made for results on the Head Start Measures Battery, but findings were reported in raw scores with no intrinsic meaning.

Source: Laurna Champ, Developmental Progress of Children Enrolled in Oklahoma Head Start Programs in 1981-1988, unpublished manuscript, 1988.

Does Head Start Make a Difference?

Authors: Janet Currie and Duncan Thomas

Outcome area studied: Cognitive and health

Overview of study: Examined the impact of Head Start on school performance, cognitive attainment, and various health and nutritional measures

Design: Post-test only, comparison groups comprised participants in other preschool or no preschool

Population: U.S. Head Start participants

Sample: National sample of data for nearly 5,000 children from the National Longitudinal Survey of Youth and the National Longitudinal Survey's Child-Mother file

Head Start program year(s): 1986-90

Measures/instrumentation: Picture Peabody Vocabulary Test and grade retention

Findings: Head Start had positive and persistent effects on test scores and school attainment of white children relative to participation in either other preschool or no preschool after controlling for family and background effects. An increase in test scores was noted for African American children, but these gains were quickly lost, and there appeared to be no positive effects in school attainment. Greater access to preventive health care was reported for white and African American children who attended Head Start or other preschools.

Source: Janet Currie and Duncan Thomas, "Does Head Start Make a Difference?" The American Economic Review, Vol. 85, No. 3 (1995), pp. 341-64.

A Comparison of Head Start and Non-Head Start Reading Readiness Scores of Low-Income Kindergarten Children of Guam

Author: Maria D. Esteban

Outcome area studied: Cognitive

Overview of study: Followed up Head Start participants in kindergarten

Design: Post-test only, comparison group comprised low-income kindergarten students that did not attend Head Start

Population: Head Start participants from six public schools on Guam

Sample: 35 male and 35 female Head Start children and 35 male and 35 female non-Head Start children

Head Start program year(s): 1985-86

Measures/instrumentation: Brigance K&I Screen for Kindergarten

Findings: Differences among the four groups were not significant at the $p = .05$ level. Head Start to non-Head Start comparison was not significant.

Source: Maria D. Esteban, A Comparison of Head Start and Non-Head Start Reading Readiness Scores of Low-Income Kindergarten Children of Guam, UMI Dissertation Services (Ann Arbor, Mich.: 1987).

The Effectiveness of Family Health Care in Head Start: The Role of Parental Involvement

Author: Barbara A. Facchini

Outcome area studied: Health

Overview of study: Relationship of the amount of parental involvement in the Head Start program to the amount of health care received by both Head Start-age children and their siblings

Design: Post-test only, comparison group selected from waiting list

Population: West Haven, Connecticut, Head Start program

Sample: 40 Head Start children and 20 waiting-list children for comparison group

Head Start program year(s): 1980-81

Measures/instrumentation: Immunizations, physical examinations, health screenings, and dental examinations

Findings: Immunizations were up to date for about one-half of both the Head Start children and the waiting-list children before the beginning of the Head Start programs. All of the Head Start children were up to date during the Head Start year, but only a few additional waiting-list children were up to date. Head Start children were more likely to receive health screenings and dental examinations ($p < .001$). Head Start children were more likely to receive physical examinations, but only the difference for children of highly involved parents was significantly different from the waiting-list children ($p < .05$). No significant difference for immunizations was found between the siblings of Head Start children and siblings of waiting-list children. Head Start siblings were more likely to have received health and dental screenings ($p < .05$).

Source: Barbara A. Facchini, The Effectiveness of Family Health Care in Head Start: The Role of Parental Involvement, Quinnipiac College (West Haven, Conn.: 1985).

The Effects of Head Start Health Services: Executive Summary of the Head Start Health Evaluation

Authors: Linda B. Fosburg and Bernard Brown

Outcome area studied: Health

Overview of study: Longitudinal study of the Head Start health services

Design: Pretest/post-test, longitudinal experimental design, involving random assignment of children to a Head Start and a non-Head Start group

Population: Four large Head Start programs

Sample: 208 children completed both pre- and post-tests, 609 received post-tests only

Head Start program year(s): 1980-81

Measures/instrumentation: Pediatric, dental, anthropometric, hematology, developmental, speech and language, vision, and hearing evaluations and nutritional observation; parent interview addressed the health history of child, nutritional evaluation of child, and family background

Findings: Head Start children were more likely to receive preventive and remedial health services than other low-income children in their community. Head Start children were more likely to receive medical and dental examinations, speech evaluation and therapy services, and vision screen or examination. Head Start children tested at both pretest and post-test were less likely to have speech and language deficiencies at post-test. Nutritional intake evaluation showed exceptionally positive impacts of Head Start's nutrition services on children and their families.

Source: Linda B. Fosburg and Bernard Brown, The Effects of Head Start Health Services: Executive Summary of the Head Start Evaluation, Abt Associates (Cambridge, Mass.: 1984).

Children Are a Wonderful Investment: A Study in Preschool Education

Authors: Mary Fulbright and others

Outcome area studied: Cognitive and family

Overview of study: To examine the effects of preschool education on children of low-income families in Dallas, Texas

Design: Post-test only, followed up students who had attended Sunnyview Head Start Center during the previous 5 years; comparison group selected from children in district who did not attend Sunnyview Head Start Center but were matched on demographic characteristics

Population: Dallas Independent School District students who had attended Sunnyview Head Start Center during the previous 5 years

Sample: 83 former Sunnyview parents and 76 comparison group parents; for the grade retention analysis, 43 Sunnyview and 41 comparison students

Head Start program year(s): 1984-89 (estimated)

Measures/instrumentation: Demographic, economic, home environment, and educational experience/expectation information was

collected from parents. Grade retention information was gathered from school files.

Findings: Significantly fewer former Sunnyview students had repeated a grade than comparison group students. Sunnyview parents reported significantly more educational items in the home than comparison group parents.

Source: Mary Fulbright and others, Children Are a Wonderful Investment: A Study in Preschool Education, Community Services Development Center, Graduate School of Social Work, University of Texas (Arlington, Tex.: 1989).

Health Services and Head Start: A Forgotten Formula

Authors: Barbara A. Hale, Victoria Seitz, and Edward Zigler

Outcome area studied: Health

Overview of study: Studied the impact of Head Start's health services on children and their siblings

Design: Post-test only, comparison groups from the waiting list and from a nursery school serving middle-class families

Population: Head Start participants in two adjacent small cities in Connecticut

Sample: 40 Head Start children, 18 children on the Head Start waiting list, 20 children enrolled in a nursery school, and 103 siblings of the nursery school children

Head Start program year(s): 1984-85

Measures/instrumentation: Immunizations, physical examinations, health screenings, and dental examinations

Findings: Head Start children received more age-appropriate health screenings than middle-class children and waiting-list children. Head Start children were more likely to receive dental examinations than middle-class children and waiting-list children. Head Start siblings were less likely than middle-class siblings to receive age-appropriate immunizations and health screenings.

Source: Barbara A. Hale, Victoria Seitz, and Edward Zigler, "Health Services and Head Start: A Forgotten Formula," Journal of Applied Developmental Psychology, Vol. 11 (1990), pp. 447-58.

An Analysis of the Effectiveness of Head Start and of the Performance of a Low-Income Population in MCPS

Author: Kathleen Hebbeler

Outcome area studied: Cognitive

Overview of study: Examined the long-term effectiveness of Head Start by comparing the performance of Head Start graduates in elementary and secondary school to that of students who had applied for Head Start but did not attend

Design: Post-test only, followed up three cohorts of Head Start graduates; one cohort, the 1978-79 group, was within the scope of our study; comparison group selected from waiting list for each respective year

Population: Children continuously enrolled in Montgomery County, Maryland, school system between 1980 and 1984 and currently in the fourth grade

Sample: Head Start group comprised 411 children; the comparison group had 89.

Head Start program year(s): 1978-79

Measures/instrumentation: California Achievement Test, Cognitive Abilities Test, special education placements, and grade retention

Findings: Head Start group had a higher percentage of students who scored above the 80th percentile on one of the subtests of the Cognitive Abilities Test administered in the third grade.

Source: Kathleen Hebbeler, An Analysis of the Effectiveness of Head Start and of the Performance of a Low-Income Population in MCPS, unpublished manuscript, Education Resources Information Center (ED281674), 1985.

A Comparison of the Academic Achievement of Urban Second Grade Pupils With Different Forms of Public Preschool Experience

Author: Elva Williams Hunt

Outcome area studied: Cognitive

Overview of study: Comparison of the academic achievement of urban second grade students from low-income families with different forms of public preschool experience

Design: Post-test only, followed up Head Start participants in second grade; comparison groups comprised students with public preschool (First Step) or no preschool experience

Population: Three cohorts of second grade students from the Newport News Public Schools

Sample: 74 former Head Start students, 92 former First Step preschool students, and 92 students with no preschool experience

Head Start program year(s): 1980-81, 1981-82, 1982-83

Measures/instrumentation: Standardized test scores and grade retention

Findings: Achievement test scores of the three groups were not significantly different. No conclusion was reached about the performance of Head Start students on the grade retention measure.

Source: Elva Williams Hunt, A Comparison of the Academic Achievement of Urban Second Grade Pupils With Different Forms of Public Preschool Experience, UMI Dissertation Services (Ann Arbor, Mich.: 1987).

A Head Start Program Evaluation in Terms of Family Stress and Affect: A Pilot Study

Authors: Ron Iverson and others

Outcome area studied: Family

Overview of study: Assessment of the effect of a local Minnesota Head Start's family services on family stress levels

Design: Pretest/post-test, comparison groups selected from waiting-list families and from the local population

Population: Families with children enrolled in the Clay-Wilkin Opportunity Council Head Start program

Sample: 149 Head Start families were surveyed at the beginning of the program year and completed a post-test in May and a 1-year follow-up the following May. Twenty-one waiting-list families and 35 randomly selected families with young children from the general population were surveyed as comparison groups.

Head Start program year(s): 1991-92, 1992-93

Measures/instrumentation: Index of Family Stress and Adjustment—Stress was measured by a correlated subscale of 13 of the original 55 stress items, and affect was measured by a correlated subscale of 9 of the original 30 affect items.

Findings: Head Start and waiting-list families were both significantly higher in stress means and lower in affect means than the general population families. Head Start families were both significantly lower in stress and significantly higher in affect than the waiting-list families at Head Start post-test. Head Start gains in both stress and affect measures appeared to reverse at the 1-year follow-up, but the changes were not significant.

Source: Ron Iverson and others, A Head Start Program Evaluation in Terms of Family Stress and Affect: A Pilot Study, State of Minnesota Department of Jobs and Training (Moorhead, Minn.: 1993).

Final Report-The Head Start Family Impact Project

Authors: Robert K. Leik and Mary Anne Chalkley

Outcome area studied: Family

Overview of study: Studied family functioning and optimal involvement of parents in Head Start

Design: Pretest/post-test of two treatment groups—regular Head Start and an enriched program with a comparison group selected from the Head Start waiting list

Population: Head Start participants in Minneapolis, Minnesota

Sample: 51 families in regular Head Start, 30 families in the enriched program, and 21 waiting-list families

Head Start program year(s): 1986-87

Measures/instrumentation: Various measures of family characteristics, mother's evaluation of her child's behavior and competence, and children's feelings of competence and social acceptance

Findings: Head Start families exhibited large and significant changes in family cohesion and adaptability. Mothers in both Head Start groups increased their evaluation of their children's competence. Children in all samples increased their sense of competence and acceptance.

Source: Robert K. Leik and Mary Anne Chalkley, Final Report—The Head Start Family Impact Project.

A Longitudinal Study to Determine the Effects of Head Start Participation on Reading Achievement in Grades Kindergarten Through Six in Troy Public Schools

Author: Paula J. Nystrom

Outcome area studied: Cognitive

Overview of study: Review impact of the Head Start program on academic achievement of children

Design: Post-test only, followed up Head Start participants in kindergarten through sixth grade; comparison group comprised children who had not attended Head Start, but who were similar to the Head Start group on certain demographic variables

Population: Head Start participants from three schools in Troy, Michigan

Sample: 54 Head Start children and 54 comparison children

Head Start program year(s): 1980-81, 1981-82, 1982-83, 1983-84

Measures/instrumentation: Metropolitan Readiness Test, Gates-MacGinite Reading Test, Iowa Tests of Basic Skills, Cognitive Abilities Test, special education placements, and grade retention

Findings: Mean scores at kindergarten were higher for the Head Start group compared with the comparison group at kindergarten; no significant

differences were found at any of the other grade levels. An analysis of change in scores over time showed a significant difference in favor of the Head Start group. No significant differences were found between the Head Start and comparison group for special education placement and grade retention.

Source: Paula J. Nystrom, A Longitudinal Study to Determine the Effects of Head Start Participation on Reading Achievement in Grades Kindergarten Through Six in Troy Public Schools, UMI Dissertation Services (Ann Arbor, Mich.: 1988).

A Comparison of Long Range Effects of Participation in Project Head Start and Impact of Three Differing Delivery Models

Author: Yvonne B. Reedy

Outcome area studied: Cognitive, socioemotional, and family

Overview of study: Investigated possible differences among groups of children receiving Head Start through three different delivery models

Design: Post-test only, followed up Head Start participants after 2 to 4 years in public schools to examine the long-range effects of different delivery models; comparison group comprised children who might have attended Head Start but did not

Population: Head Start participants in rural Pennsylvania

Sample: 18 were children for each of the three groups: classroom, mixed model, and home based; and 18 children were included in the control group

Head Start program year(s): Not specified

Measures/instrumentation: Woodcock-Johnson Psychoeducational Battery - Part II, Tests of Achievement; PPVT-R; Child Behavior Checklist - Parent Rating Scale; Child Behavior Checklist - Teacher Rating Scale; Vineland Adaptive Behavior Scale - Survey Form; and Head Start Follow-up Family Questionnaire

Findings: No differences among Head Start and non-Head Start children in reading, math, written language, or receptive language. Levels were in the average range when compared with national norms. Head Start children obtained significantly higher mean scores on the measure of

general knowledge. Head Start children had significantly lower mean scores on both subscales and total scale on the measure of maladaptive behavior. Correlations with teacher reports were significant. On the socialization scale, differences were not significant at $p = .05$. On the adaptive behavior measures, the non-Head Start children obtained significantly higher means on the communication, daily living skills, and social skills domains, as well as on the total adaptive behavior score.

On the parent questionnaire, non-Head Start parents reported they felt less capable of providing a good learning environment, spent less time working with the child on homework or other learning activities, were less likely to seek information about age-appropriate expectations, were more likely to resort to spanking as a form of discipline, were less able to find community services when needed and to feel their involvement with their child's education had resulted in any noticeable accomplishments. On the daily living skills, social skills, and total independent living scales, the children in the classroom model obtained lower means than the two groups who received home visits. Parents of children in the classroom model reported they spent smaller amounts of time working with their children at home, and they were less likely to seek out information about age-appropriate information and to feel that their involvement in their children's education resulted in any noticeable accomplishments.

Source: Yvonne B. Reedy, A Comparison of Long Range Effects of Participation in Project Head Start and Impact of Three Differing Delivery Models, Pennsylvania State University (State College, Penn.: 1991).

A Study of Duration in Head Start and Its Impact on Second Graders' Cognitive Skills

Author: Joyce Harris Roberts

Outcome area studied: Cognitive and socioemotional

Overview of study: Assessed the impact of Head Start programming on later school success and the development of social competence in its graduates

Design: Post-test only, compared 1-year Head Start participants with 2-year Head Start participants and a non-preschool comparison group of second grade classmates

Population: Second grade students in four public schools in a large suburban school district

Sample: 30 children with 1 year of Head Start year, 22 children with 2 years of Head Start, and 33 children with no preschool

Head Start program year(s): 1978-79, 1979-80

Measures/instrumentation: Locus of Control Scale for Children - Pre-School and Primary, Form A; Self-Concept Inventory; and Cognitive Abilities Test (Primary Level)

Findings: No significant difference was found between groups on any measure.

Source: Joyce Harris Roberts, A Study of Duration in Head Start and Its Impact on Second Graders' Cognitive Skills, UMI Dissertation Services (Ann Arbor, Mich.: 1984).

Changes in Mental Age, Self-Concept, and Creative Thinking in Ethnically Different 3- and 4-Year-Old Head Start Students

Author: Linda L.B. Spigner

Outcome area studied: Cognitive

Overview of study: Studied Head Start participants' progress after 8 months of Head Start participation and conducted home interviews for the children who showed the highest gains and for the children who made the least progress

Design: Pretest/post-test

Population: Head Start participants in a north Texas community

Sample: 37 Head Start participants

Head Start program year(s): Exact year not specified

Measures/Instrumentation: Bankson Language Screening Test, Developmental Test of Visual-Motor Integration, Peabody Picture Vocabulary Test, Self-Concept Adjective Checklist, and Torrance Tests of Creative Thinking

Findings: Average mental age gain of almost 11 months was significant at the .01 level. Gains in self-concept and creative thinking were significant at the .01 level.

Source: Linda L.B. Spigner, Changes in Mental Age, Self-Concept, and Creative Thinking in Ethnically Different 3- and 4-Year-Old Head Start Students, Texas Woman's University (Denton, Tex.: 1985).

Learning by Leaps & Bounds

Author: Texas Instruments Foundation, Head Start of Greater Dallas, and Southern Methodist University

Outcome area studied: Cognitive

Overview of study: Followed up Margaret H. Cone Preschool Head Start program participants; cohorts 4, 5, and 6 participated in a new Language Enrichment Activities Program

Design: Post-test only, comparison group comprised classmates who did not attend the Margaret H. Cone Preschool

Population: Six cohorts of children attending the Margaret H. Cone Preschool Head Start program in Dallas, Texas

Sample: Cohorts ranged from about 30 to 58 children

Head Start program year(s): 1990-96

Measures/instrumentation: Battelle Developmental Inventory, PPVT-R, Clinical Evaluation of Language Fundamentals - Preschool, and Iowa Test of Basic Skills

Findings: Results of cohorts 2, 3, 4, 5, and 6 revealed a pattern of improved performance in vocabulary, language skills, concept development, and social-adaptive skills during the years of the language enrichment program.

Source: Texas Instruments, Head Start of Greater Dallas, and Southern Methodist University, Learning by Leaps & Bounds, paper presented at Head Start's Third National Research Conference (Washington, D.C.: June 22, 1996).

Early Childhood Educational Intervention: An Analysis of Nicholas County, Kentucky, Head Start Program Impacts From 1974-1986

Author: Marium T. Williams

Outcome area studied: Cognitive

Overview of study: Examined the impact of the Nicholas County Head Start Program over a 12-year period

Design: Post-test only, followed up Head Start participants in first grade through sixth grade; comparison groups were selected from comparable first grade enrollment

Population: Children who entered first grade in the years 1975, 1976, 1979, 1980, and 1981 in Nicholas County, Kentucky; the first three groups are outside our period of study

Sample: 14 Head Start and 9 comparison children for 1979-80 and 11 Head Start and 10 comparison children for 1980-81

Head Start program year(s): 1979-80, 1980-81

Measures/instrumentation: Comprehensive Tests of Basic Skills, Cognitive Skills Index, mathematics and reading/English grades, Kentucky Essential Skills Test, special education placements, and grade retention

Findings: No significant differences were found for most comparisons. Reading scores for the Head Start children were significantly better than those of the comparison group for 3 of the 6 years at the .05 level of significance.

Source: Marium T. Williams, Early Childhood Educational Intervention: An Analysis of Nicholas County, Kentucky, Head Start Program Impacts From 1974-1986, UMI Dissertation Services (Ann Arbor, Mich.: 1988).

Is an Intervention Program Necessary in Order to Improve Economically Disadvantaged Children's IQ Scores?

Authors: Edward Zigler and others

Outcome area studied: Cognitive

Overview of study: Studied changes in intelligence quotient scores of children attending Head Start

Design: Pretest/post-test, comparison group comprised Head Start-eligible children not attending Head Start; testing was done at three points in the Head Start year

Population: Preschool children from economically disadvantaged families living in low-income, inner-city neighborhoods in New Haven, Connecticut

Sample: 59 Head Start children and 25 comparison children

Head Start program year(s): Not stated

Measures/instrumentation: Stanford-Binet Intelligence Scale, Form L-M

Findings: Both groups increased from test to retest, a result which was attributed to familiarity with the testing situation. Only the Head Start group continued to show improvement on the post-test, which was interpreted as reflecting changes in the children's motivation from attending a preschool intervention program.

Source: Edward Zigler and others, "Is an Intervention Program Necessary in Order to Improve Economically Disadvantaged Children's IQ Scores?" Child Development, Vol. 53 (1982), pp. 340-48.

Comments From the Department of Health and Human Services

DEPARTMENT OF HEALTH & HUMAN SERVICES

Office of Inspector General

Washington, D.C. 20201

MAR 11 1997

Ms. Carlotta C. Joyner
Director, Education and
Employment Issues
United States General
Accounting Office
Washington, D.C. 20548

Dear Ms. Joyner:

Enclosed are the Department's comments on your draft report, "Head Start: Existing Research Provides Little Information on Impact of Modern-Day Program." The comments represent the tentative position of the Department and are subject to reevaluation when the final version of this report is received.

The Department appreciates the opportunity to comment on this draft report before its publication.

Sincerely,

June Gibbs Brown
Inspector General

Enclosure

The Office of Inspector General (OIG) is transmitting the Department's response to this draft report in our capacity as the Department's designated focal point and coordinator for General Accounting Office reports. The OIG has not conducted an independent assessment of these comments and therefore expresses no opinion on them.

BEST COPY AVAILABLE

Comments of the Department of Health and Human Services on the U.S. General Accounting Office's Draft Report, "Head Start: Existing Research Provides Little Information on Impact of Modern-Day Program," GAO/HEHS-97-59

General Comments

We appreciate the opportunity to comment on this report and wish to commend GAO on its collaborative approach to designing and conducting this project. Our response contends that the research base on the efficacy of Head Start is more substantial than depicted in the GAO report, and that the strategy that the Department has been following to extend this base is carefully thought out and designed to produce important findings on the impacts and most important active ingredients in Head Start program quality.

GAO Recommendation

We recommend that HHS include in its research plan studies that assess the impact of regular Head Start programs.

Department Comments

1. There is clear evidence of the positive impacts of Head Start services.

There is a large body of convincing research on the short and intermediate effects of Head Start for low-income children and their families. Head Start helps children get ready and stay ready for school and supports parents in improving their parenting skills, enhancing their abilities to contribute to their communities, and, in many cases, accelerates their progress towards economic self-sufficiency.

Specifically, evaluations of Head Start programs in the 1970s and early 1980s provided important context for the research and evaluation plans of the 1990s. In 1985, McKey et al., conducted a synthesis of Head Start research and evaluation studies that had been conducted up to that time. Because most of the studies were conducted of children attending Head Start prior to 1976, GAO contends that their value should be dismissed. However, the synthesis study findings are based on a much larger body of work than the GAO report, including 72 studies of cognitive effects.

- The 21 most rigorous studies (those with experimental treatment-control group designs), showed an average effect size of .52 across all cognitive measures.

- Across 17 studies, Head Start participation resulted in educationally meaningful effect sizes on achievement motivation and social behavior. These findings were robust enough to appear across 18 different measures of achievement motivation, and 11 measures of social behavior.

- The examination of 34 studies on health found positive Head Start effects on child health.

- A review of 75 studies on Head Start's effects on families found benefits including positive effects on maternal well-being and depression as a result of Head Start participation.

- Twenty-nine studies examined the effects of Head Start on communities and found positive effects on improvement in community services and service to the poor.

Even within the time parameters set out by GAO, there is evidence of Head Start's effectiveness in multiple domains. The GAO report includes the Effects of Head Start Health Services study (cited as Fosburg and Brown) in Appendix III, and although the study meets GAO's requirements for an impact evaluation, GAO effectively dismisses its findings. The study was a longitudinal study with over 800 subjects randomly assigned to Head Start or a control group. The study found positive effects on nutrient intake, nutrient quality in meals served at home, health improvements in pediatric, dental, developmental, speech, language, and vision problems. The effects of the program were especially pronounced in communities where access to health care is difficult and medical care was poor.

The recent Currie and Thomas (1995) analysis of a national sample of Head Start children from the National Longitudinal Survey of Youth (NLSY) (also cited in Appendix III) found that, controlling for family background, not only did Head Start children perform better than siblings who attended no preschool, but also (for white children), the effects of Head Start were much greater than the effects of attending other preschool programs. For African-American children, initial increases in cognitive test scores were observed but longer-term effects were not detected. Follow-up research is assessing the extent to which African-American children experience more adverse schooling conditions after the Head Start year. Currie and Thomas also found that participation in Head Start was associated with an 8 to 9 percent higher probability of being immunized.

2. Based on the recommendations of leading researchers, the Department is conducting a well-balanced, innovative set of new studies of Head Start, including a significant series of longitudinal research efforts.

The research agenda that Head Start has been following since 1990 is one that was designed in consultation with and at the recommendation of the foremost researchers, child development experts and policy analysts in the country. It was designed to provide the information necessary to improve the effectiveness of Head Start in an environment where the characteristics of the population have changed considerably and where risk factors which threaten optimal development of poor children have multiplied.

In the past few years, Head Start has also placed major research emphasis on quality, since a large body of research has demonstrated the relationship between quality programs and program effects.

The Department contends that GAO's description of Head Start's research plan does not acknowledge major longitudinal studies that will produce information on aspects of Head Start effects and the variables that contribute to or detract from the maintenance of those effects.

- The Department of Education's Early Childhood Longitudinal Study will follow a nationally representative sample of 25,000 kindergarten children, including a substantial subsample of children who attended Head Start.

- In a partnership with the National Institute of Child Health and Human Development, Head Start is contributing to the Study of Early Child Care, a 7-year longitudinal study of over 1,200 infants and families from 10 sites across the United States. Children and families will be followed as they move through a variety of child care, Head Start, and other informal and program environments where cognitive, health and social-emotional outcomes of children will be studied.

- Head Start is also working with the National Institute of Mental Health on a longitudinal study of needs, uses, costs and outcomes of mental health services for young children and adolescents, including an oversample of Head Start eligible children.

- Additional valuable data are in the Head Start/Public School Transition Demonstration Study, where children in both the experimental and control groups were in regular Head Start classrooms and from kindergarten on, there are data on children who did not attend Head Start.

In addition, the newly launched Early Head Start Evaluation (EHS) is both a longitudinal study and a major new experimental study of impacts. The Department's selection of an experimental impact design for this evaluation took into account the new nature of these services, the extent to which programs would not be able to serve all eligible children and could establish randomly assigned control groups, and other feasibility issues. The evaluation includes a large sample of families in sites that were carefully selected to reflect the diversity in geography, community, population characteristics, and program approaches that exist in regular Head Start in the 1990s. The EHS evaluation also includes an extensive implementation study and assessment of services received by control-group families so that any treatment-control differences can be more accurately interpreted. The study will also provide correlational data on a number of configurations of services, e.g., Early Head Start plus regular

Head Start, Early Head Start plus some other program, Head Start only for control group families, and other infant and toddler child care programs plus Head Start.

3. The Department is building a substantial system of innovative research, development and management tools in response to the Government Performance and Results Act.

The GAO states that random assignment studies on Head Start effects are needed in light of the new results-oriented environment of the Federal Government. However, the Department has created a wealth of new initiatives and systems to promote stronger accountability and program quality, notably the Head Start Quality Research Consortium. In 1995, the Head Start Bureau established four Quality Research Centers (QRCs). One objective of the QRCs is to work cooperatively to develop and test instruments to gather information on program performance measures. In 1996, the Head Start Bureau established a Performance Measure Center which will use these instruments to conduct the Head Start Family and Child Experiences Survey (FACES). The FACES is an ambitious effort to assess the performance of the Head Start program on an ongoing basis through a national longitudinal study of a representative sample of Head Start children and their families. The work that is now in progress on performance measures is an essential element in a results oriented approach and has the additional benefit of establishing a baseline for Head Start from which improvements in outcomes can be measured on an ongoing basis.

The Head Start FACES project will provide valuable information on the overall effectiveness of Head Start and the relative value of specific approaches and procedures being implemented by different grantees across the United States. Although the conclusions of these research efforts may not be as rigorous and unarguable as studies carried out with random assignment experimental designs, they will permit some inferences to be made about causal mechanisms. The use of advanced multivariate analytic techniques, such as hierarchical modeling, in the quality research and performance measures studies will permit statistical control of potentially confounding factors.

In addition, the four QRCs are currently evaluating possible strategies for forming comparison groups of children from low-income families who are eligible for Head Start, but who do not receive it because of lack of sufficient slots, parent choice, or other reasons. The QRCs are also evaluating the feasibility of using random assignment to determine which of these eligible children receive Head Start services in a given year. Once the QRCs have come up with workable strategies for case-control matching or random assignment, they will carry out an integrated series of relatively small-scale evaluation studies that will address important impact questions. Furthermore, once the viability of specific evaluation study designs has been

demonstrated at the local level, the Department may consider implementing such a study or series of studies on a larger scale.

A final issue which is overlooked in the GAO report is the importance of studying the quality of Head Start and other early childhood programs along with assessing program impacts. It is only through a better understanding of what quality entails and how program components and features contribute to child and family outcomes that we will be able to design and implement programs that maximize benefits while keeping costs reasonable. A better understanding of quality will give Federal, State, and local policymakers more options to choose from in tailoring early childhood programs to meet the needs and resources of particular communities and particular groups of children. It will also give them a better appreciation of the probable consequences of their choices for the development, learning and behavior of children from low-income families.

Acknowledgments

Many researchers and early childhood experts provided valuable assistance and information used in producing this report. In particular, we wish to acknowledge the following individuals who reviewed the draft report: Dr. Richard Light, Harvard University; Dr. Mark Lipsey, Vanderbilt University; Greg Powell, National Head Start Association; and Dr. Edward Zigler, Yale University. Although these reviewers provided valuable comments, they do not necessarily endorse the positions taken in the report.

Appendix VI

GAO Contacts and Staff Acknowledgments

GAO Contacts

D. Catherine Baltzell, Assistant Director, (202) 512-8001
Deborah L. Edwards, Evaluator-in-Charge, (202) 512-5416

Staff Acknowledgments

In addition to those named above, the following individuals made important contributions to this report: Sherri Doughty managed the literature search and co-wrote the report, Wayne Dow led the literature review and screening, and Paula DeRoy performed the literature searches and collected the manuscripts.

Related GAO Products

Executive Guide: Effectively Implementing the Government Performance and Results Act (GAO/GGD-96-118, June 1996).

Head Start: Information on Federal Funds Unspent by Program Grantees (GAO/HEHS-96-64, Dec. 29, 1995).

Early Childhood Centers: Services to Prepare Children for School Often Limited (GAO/HEHS-95-21, Mar. 21, 1995).

Program Evaluation: Improving the Flow of Information to the Congress (GAO/PEMD-95-1, Jan. 30, 1995).

Early Childhood Programs: Parent Education and Income Best Predict Participation (GAO/HEHS-95-47, Dec. 28, 1994).

Early Childhood Programs: Local Perspectives on Barriers to Providing Head Start Services (GAO/HEHS-95-8, Dec. 21, 1994).

Early Childhood Programs: Multiple Programs and Overlapping Target Groups (GAO/HEHS-95-4FS, Oct. 31, 1994).

Early Childhood Programs: Many Poor Children and Strained Resources Challenge Head Start (GAO/HEHS-94-169BR, May 17, 1994).

Infants and Toddlers: Dramatic Increase in Numbers Living in Poverty (GAO/HEHS-94-74, Apr. 7, 1994).

Poor Preschool-Age Children: Numbers Increase but Most Not in Preschool (GAO/HRD-93-111BR, July 21, 1993).

Ordering Information

The first copy of each GAO report and testimony is free. Additional copies are $2 each. Orders should be sent to the following address, accompanied by a check or money order made out to the Superintendent of Documents, when necessary. VISA and MasterCard credit cards are accepted, also. Orders for 100 or more copies to be mailed to a single address are discounted 25 percent.

Orders by mail:

U.S. General Accounting Office
P.O. Box 6015
Gaithersburg, MD 20884-6015

or visit:

Room 1100
700 4th St. NW (corner of 4th and G Sts. NW)
U.S. General Accounting Office
Washington, DC

Orders may also be placed by calling (202) 512-6000 or by using fax number (301) 258-4066, or TDD (301) 413-0006.

Each day, GAO issues a list of newly available reports and testimony. To receive facsimile copies of the daily list or any list from the past 30 days, please call (202) 512-6000 using a touchtone phone. A recorded menu will provide information on how to obtain these lists.

For information on how to access GAO reports on the INTERNET, send an e-mail message with "info" in the body to:

info@www.gao.gov

or visit GAO's World Wide Web Home Page at:

http://www.gao.gov

PRINTED ON RECYCLED PAPER

BEST COPY AVAILABLE

United States
General Accounting Office
Washington, D.C. 20548-0001

Bulk Rate
Postage & Fees Paid
GAO
Permit No. G100

Official Business
Penalty for Private Use $300

Address Correction Requested

BEST COPY AVAILABLE